The FHWA Travel Model Improvement Program Workshop over the Web

The Travel Model Development Series: Part I – Travel Model Estimation

presented by
Thomas Rossi
Yasasvi Popuri
Cambridge Systematics, Inc.

February 10, 2009

1

Webinar Outline

- Session 1: Introduction – October 16, 2008
- Session 2: Data Set Preparation – November 6, 2008
- Session 3: Estimation of Non-Logit Models – December 11, 2008
- Session 4: Estimation of Logit Models – February 10, 2009

2

Webinar Outline (continued)

- Session 5: Application and Validation of Logit Models – March 12, 2009
- Session 6: Advanced Topics in Discrete Choice Models – April 14, 2009
- Session 7: Trip Assignment – May 7, 2009
- Session 8: Evaluation of Validation Results – June 9, 2009

3

Homework
From Session 3

4

The Use of Logit Models in Transportation Planning

- **Can be used to analyze any choice made by travelers with discrete alternatives**
- **Mode choice is the most common application for which logit models are used in transportation planning**
- **But there are many other choice processes for which logit models serve well**

5

How Logit Models Work

- Define set of alternatives for the choice
- Define and compute the “utility” of each alternative for each choice maker
- Compute probability of choosing each alternative
- Determine choice share for each alternative

6

Utility

- Defined as the attractiveness of an alternative based on its attributes and other variables
- Expressed as a function (usually linear)

7

Formulation of the Random Utility Model

$$U_i = V_i + e_i$$

where:

U_i = Utility of alternative i

V_i = Deterministic component of U_i

e_i = Random component of U_i

P(i) = P [$U_i > U_j$ for all *j* in the choice set]

8

Utility Function

$$U_i = B_{0i} + B_{1i} X_{1i} + B_{2i} X_{2i} + \cdots + B_{ni} X_{ni} + e_i$$

where:

B_{ki} = coefficient for variable X_{ki} for alternative i

X_{ki} = variable that explains choice for alternative i

9

Alternative Specific Constants

- Intended to estimate the effects of things that might favor one alternative over another but cannot be quantified
- Assume one constant per alternative, with one assumed to be zero (called the "base alternative")

10

Indicator ("Dummy") Variables

- Have one of two values, zero or one
- Indicate whether characteristic of traveler, trip, or context is true
- Examples: Income level of household, vehicle availability, peak period

11

Example
Mode Choice Model

- Alternatives are modes (typically 2 to ~25)
- Utilities represent characteristics of modes, travelers, and context
- Choice probabilities used to determine mode shares

12

Example
Destination Choice (Trip Distribution)

- Alternatives are zones
- Utilities represent travel impedance and characteristics of destinations, travelers, and context
- Choice probabilities used to determine shares of trips to each zone

13

Example
Vehicle Availability Model

- Alternatives are number of vehicles owned (0, 1, 2, 3, etc.)
- Utilities represent characteristics of households, areas, and context
- Choice probabilities used to determine shares of households for each vehicle availability level by zone

14

Other Examples

- Time of day choice models
- Socioeconomic models
- Trip generation
- Land use models
- Components of activity based models

15

The Binary Logit Formulation

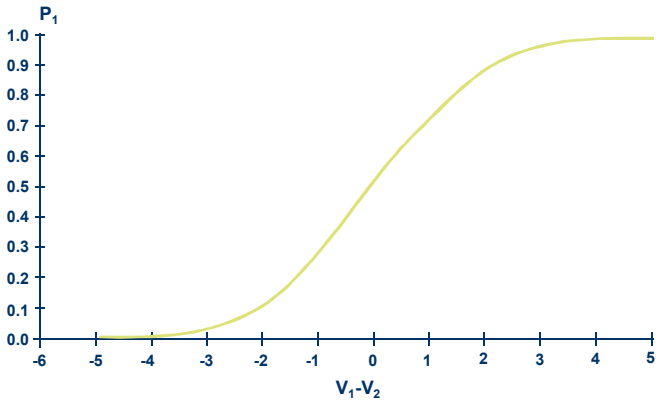
$$P_r(1) = \frac{\exp(v_1)}{\exp(v_1) + \exp(v_2)}$$

$$P_r(1) = \frac{1}{1 + \exp[-(v_1 - v_2)]}$$

Only the differences in utilities matter

16

The Binary Logit Model



17

The Multinomial Logit Model

Extending the binary model to three modes:

$$P(1) = \frac{\exp(v_1)}{\exp(v_1) + \exp(v_2) + \exp(v_3)}$$

Extending to n modes:

$$P(1) = \frac{\exp(v_1)}{\exp(v_1) + \exp(v_2) + \cdots + \exp(v_n)}$$

18

MNL Model Properties

- P_i, all i depends on the deterministic components of the utilities of all alternatives (V_j, all j)
- P_i increases as V_i increases, and decreases as V_j ($j \neq i$) increases
- P_i depends only on the values ($V_j - V_i$) for all alternatives j, except i

19

The IIA Property
Definition

- The independence from irrelevant alternatives property

  "For any individual, the ratio of the probabilities of choosing two (available) alternatives is independent of the availability or attributes of any other alternative"

- Mathematically

$$\frac{\Pr(i)}{\Pr(k)} = \frac{\exp(V_i)}{\exp(V_k)} = \exp(V_i - V_k)$$

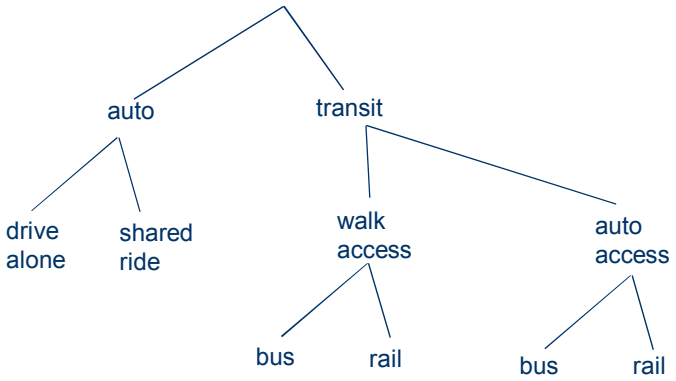
20

The IIA Property
Red and Blue

- Scenario 1
  - Available modes are auto (da) and red buses (rb); red buses have plenty of seats for all passengers
  - $V_{da} = V_{rb}$
  - MNL model says Pr(da) = Pr(rb) = 0.5
- Scenario 2
  - A new bus operator exactly duplicates red bus service using blue buses (bb)
  - MNL model says Pr(da) = Pr(rb) = Pr(bb) = 0.33

21

The Nested Logit Model



22

Utilities in Nested Logit

- Logsum variable
  - Logsum = ln Σ [exp (V_i)] for all alts. i in the nest
- Utility for nest
  - $V_{nest} = B_{0(nest)} + B_{logsum}$ logsum
  - $0 < B_{logsum} < 1$

23

Choosing the Independent Variables in Utility Functions

- Relevance to the travel choice
- Availability in the estimation data set
- Availability for model application (forecasting)
- Statistical testing

24

Setting up Model Estimation Data Set

- Survey data organized by individual records
  - For example, for mode choice, each record is a trip
- Additional variables added to data set
  - Information on transportation level of service (from skims)
  - Information at zone level, such as parking costs (from lookup tables)
  - Transformations of variables, as needed

25

Setting up Model Estimation Process

- Define utility functions
- Prepare input files for model estimation software

26

Trip Data File
Typical Fields

From the survey

- Origin and destination
- Trip purpose
- Chosen mode
- Time of day of trip
- *Trip time/cost*
- Household/person characteristics (linked from household/person file)

From other sources

- Travel time (in-vehicle)
- Other time components (wait, access/egress, transfer)
- Costs (parking, auto operating, transit fare)
- Number of transit transfers
- Zone attributes
- Logsums from other models

27

Example Data Set
Mode Choice Model Estimation (Partial)

Trip ID	Trip Purpose	Chosen Mode	Origin	Dest	IVT	Transit Walk Access		
						Wait	Walk	Fare
10001	1	1	107	12	20.1	10	15	$1.00
10002	3	1	12	118	10.3	5	2	$1.00
10003	2	1	118	107	17.2	10	15	$1.00

28

Estimating the Logit Model

1. Run statistical software to estimate coefficients of utility functions
2. Evaluate results
3. Revise specification and reestimate
   - Consider alternative variable definitions, combinations
   - Eliminate variables as appropriate
   - Consider constraining parameter estimates, but only when necessary
4. Choose "best" specification

29

Constraining Coefficient Estimates

- Inequality constraints, e.g. $B_k \leq 0$
- Fixed value constraints, e.g. $B_k = C$
- Linear constraints, e.g. $B_k = C\ B_j$
  - Assumed value of time: $B_{ivt} = (VOT)\ B_{cost}$

30

Constraining Coefficient Estimates
Why Do It?

- Estimated values not reasonable in sign or magnitude (e.g. $B_{ivt} > 0$)
- Estimate of parameter value unreasonable relative to others (e.g. $|B_{ivt}| > |B_{ovt}|$)
- Consistency with other sources desired (e.g. FTA New Starts Guidelines)

31 31

Evaluating the Logit Model Estimation Results

1. Reasonableness of coefficient estimates
   - Sign
   - Magnitudes
2. Significance of estimates (t-statistics)
   - $| t | > 1.96$ implies significance at 95% level
3. Goodness of fit (ρ^2)

32

Vehicle Availability Model
Example Estimation Results

	Vehicle Availability Level				
Variable	0	1	2	3	4+
Persons per household	--	--	0.1164 (2.1)	0.1164 (2.1)	0.2571 (2.1)
Workers per household	--	--	0.4915 (5.2)	1.474 (10.8)	2.139 (10.0)
Household density	--	-0.0458 (-2.9)	-0.1327 (-5.4)	-0.1717 (-4.4)	-0.2549 (-3.0)
ln(income)	--	1.130 (8.7)	2.497 (13.9)	2.995 (12.7)	3.242 (7.6)
Transit/highway accessibility	--	-1.133 (-1.7)	-2.054 (-2.8)	-2.742 (-3.3)	-2.742 (-3.3)
Persons less than vehicles	--	--	-2.870 (-8.8)	-1.017 (-5.3)	-0.5181 (1.1)
Constant	--	0.164 (0.2)	-3.761 (-4.6)	-8.229 (-8.0)	-12.87 (6.8)

33

Vehicle Availability Model
Example Estimation Results

	Vehicle Availability Level				
Variable	0	1	2	3	4+
Persons per household	--	--	0.1164 (2.1)	0.1164 (2.1)	0.2571 (2.1)
Workers per household	--	--	0.4915 (5.2)	1.474 (10.8)	2.139 (10.0)
Household density	--	-0.0458 (-2.9)	-0.1327 (-5.4)	-0.1717 (-4.4)	-0.2549 (-3.0)
ln(income)	--	1.130 (8.7)	2.497 (13.9)	2.995 (12.7)	3.242 (7.6)
Transit/highway accessibility	--	-1.133 (-1.7)	-2.054 (-2.8)	-2.742 (-3.3)	-2.742 (-3.3)
Persons less than vehicles	--	--	-2.870 (-8.8)	-1.017 (-5.3)	-0.5181 (1.1)
Constant	--	0.164 (0.2)	-3.761 (-4.6)	-8.229 (-8.0)	-12.87 (6.8)

ρ^2 w.r.t zero = 0.447 ρ^2 w.r.t constants = 0.302

34

Mode Choice Model
Example Estimation Results

Variable	Drive Alone	Shared Ride	Walk	Transit Auto Access	Transit Walk Access
In-vehicle time (min)	-0.020 (-2.7)	-0.020 (-2.7)	--	-0.020 (-2.7)	-0.020 (-2.7)
Out-of-vehicle time (min)	--	--	--	-0.045 (-3.7)	-0.045 (-3.7)
Cost ($)	-0.400 (-7.0)	-0.400 (-7.0)	--	-0.400 (-7.0)	-0.400 (-7.0)
Distance (miles)	--	--	-1.63 (-7.2)	--	--
Zero vehicles owned	--	--	1.18 (1.7)	1.16 (1.8)	2.49 (4.6)
Vehicles/person	--	-2.45 (-10.0)	-6.87 (-6.2)	-5.20 (-6.8)	-5.20 (-6.8)
Population density (1000/acre)	--	--	0.042 (3.7)	0.030 (3.1)	0.030 (3.1)
Employment density (1000/acre)	--	--	0.31 (4.2)	0.079 (1.2)	0.079 (1.2)
Constant	--	1.65 (8.3)	3.75 (5.9)	3.55 (6.5)	6.83 (14.2)

ρ^2 w.r.t zero = 0.319 ρ^2 w.r.t constants = 0.156

35

Typical Selected Home-Based Work Parameters

	Average from U.S. Cities	Typical Range
In-vehicle time	-0.028	-0.01 to -0.05
Out-of-vehicle time	-0.054	-0.03 to -0.07
Cost	-0.720	-0.2 to -1.3
Ratio: B_{ovt} / B_{ivt}	1.9	1.5 to 3.0
Value of Time	$2.30	$2.00 to $5.00

36

Estimation Problems

- Use of too many alternative-specific constants
- Constraining coefficients across alternatives for non-varying variables
- Perfect collinearity of variables

37

Homework
Session 4

38